

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/18975>

Please be advised that this information was generated on 2017-12-05 and may be subject to change.

Over eiwitten, kaartlezen en genomen melken

Uitgebreide tekst van de rede
uitgesproken bij de aanvaarding van het ambt van
bijzonder hoogleraar in de Structuurchemie van Eiwitten aan de
Faculteit der Natuurwetenschappen, Wiskunde en Informatica (NWI)
vanwege de Stichting Nijmeegs Universiteitsfonds (SNUF)
aan de Katholieke Universiteit Nijmegen
op vrijdag 9 maart 2001

door

Roland J. Siezen

Katholieke Universiteit Nijmegen



Voor mijn moeder

Voormeer informatie:

R. J. Siezen:

<http://www.cmbi.kun.nl/staf/siezen.shtml>

CMBI, KUN:

<http://www.cmbi.kun.nl>

WCFS:

<http://www.wcfs.nl>

NIZO food research:

<http://www.nizo.com>

Mijnheer de Rector Magnificus,
zeer gewaardeerde toehoorders,

Velen van u hier in de zaal hebben mij in de afgelopen maanden al gevraagd: “Hoe vaak zit je daar in Nijmegen, en wat doe je daar? Moet je ook onderwijs geven? Welk onderzoek ga je daar doen? Voor wie werk je nu eigenlijk, NIZO food research, het Wageningen Centre for Food Sciences (WCFS), of voor CMBI, het Centrum voor Moleculaire en Biomoleculaire Informatica van de Universiteit Nijmegen?”

De meesten vroegen: “Heb je je rede al klaar?”

Vanmiddag zal ik proberen een aantal van deze vragen te beantwoorden. Ik begin met een hele eenvoudige schets om u het overzicht te geven. Wat is de wetenschappelijke vraag, welke kennis moeten we ontwikkelen, wat zijn de doelstellingen?

Ik ga het hebben over levende cellen. De eenvoudigste zijn bacteriën, die maar uit één cel bestaan. Ons onderzoek richt zich met name op bacteriën die belangrijk zijn bij de voedselbereiding en bij voedselbederf. Een cel is eigenlijk een minifabriekje. Het neemt allerlei voedingsstoffen op, zoals koolhydraten, eiwitten en mineralen. Binnen de cel vinden allerlei stofwisselingsprocessen plaats, er wordt energie gegenereerd en er worden bouwstenen aangemaakt. Hiermee worden diverse producten gemaakt die de bacterie nodig heeft om te groeien en te overleven, en die binnen of buiten de cel terechtkomen. Hoe dat allemaal precies in zijn werk gaat binnen bacteriële cellen is deels bekend, maar deels ook onbekend – geen *black box*, maar een soort *grey box*. Waarom kan de ene bacterie bepaalde stoffen opnemen, aanmaken of afscheiden en de andere bacterie niet? U kunt uit de titel van mijn rede misschien al raden dat *eiwitten* hierin een centrale rol spelen.

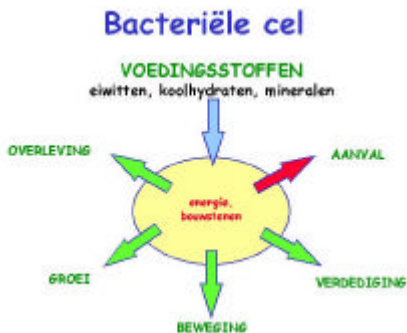
Wel, de volledige informatie voor alle eigenschappen is vastgelegd in het DNA van de bacterie, het erfelijke (of genetische) materiaal, in zijn genen. De uitdaging is dus om de volledige genetische kaart – ook wel *genoom* genoemd – van bacteriën in handen te krijgen, en vervolgens de *kaart te lezen* en te interpreteren. Hierbij zal *bio-informatica*, het gebruik van computers, mijn belangrijkste gereedschap zijn. Uiteindelijk is het doel om deze genomen te *melken*, dat wil zeggen, uit de volledige genetische kaarten kennis te ontwikkelen, voorspellen de kennis, om deze bacteriën optimaal te benutten of te bestrijden.

Ik zal nu dit proces verder toelichten, te beginnen met de *eiwitten*.

EIWITTEN

Bij het woord “eiwitten” denken de meesten van u misschien aan voedingseiwitten van plantaardige of dierlijke oorsprong, zoals tarwe-eiwitten, aardappeleiwitten of melkeiwitten. Dit noemen we “bulkeiwitten”. Zij worden meestal volledig afgebroken door stofwisselingsprocessen, denk maar aan het spijsverteringsstelsel bij ons, waardoor bouwstenen ontstaan voor de groei. Maar er zijn veel meer “specialistische eiwitten”. Alle levende organismen bevatten duizenden verschillende eiwitten die zeer bijzondere functies hebben in het lichaam, binnen cellen of juist erbuiten.

Eiwitten en hun structuur vormen de rode draad door mijn wetenschappelijke carrière. Het begon al tijdens mijn biochemiestudie aan de Rijksuniversiteit Groningen, waar ik hoofdvak- en promotie-onderzoek deed aan het eiwit *hemocyanine* van de wijngaardslak, onder leiding van



Wil Konings (toen nog promovendus) en professor Erni van Bruggen (mijn promotor). Dit bloedeiwit vervoert zuurstof, en was toen vooral bekend door zijn grote afmetingen en prachtige cilindrische structuur, die in een elektronenmicroscop goed te zien is. De afmeting van hemocyanine is ongeveer 30 nanometer; een nm is een miljoenste millimeter. De meeste eiwitten zijn nog tien keer zo klein. Naast eiwitten ga ik het hebben over cellen, die grofweg 1 à 10 micrometer groot zijn, een duizendste tot een honderdste

millimeter.

Na mijn promotie ben ik met mijn gezin twaalf jaar lang de wereld rondgetrokken, van de ene universiteit naar de andere, San Diego, Canberra in Australië, Boston en ja, zelfs Nijmegen tussendoor. Wel was het steeds onderzoek aan eiwitten maar iedere keer vanuit een andere benadering en discipline: biochemie, medische chemie, en fysische chemie. Tenslotte kwam ik terug in Nederland bij NIZO food research, waar ik eiwitten bestudeer vanuit een microbiologische, fysiologische en genetische invalshoek.

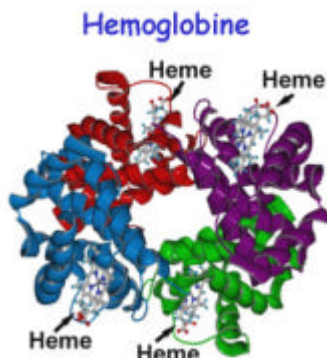
Structuurchemie van eiwitten

Eiwitten zijn biopolymeren, en hun bouwstenen zijn aminozuren. Er zijn 20 verschillende soorten aminozuren. Ze bestaan allemaal uit een identieke kopgroep (de aminogroep) en een identieke staartgroep (de zuurgroep), maar ze hebben zijgroepen die verschillen in grootte, structuur, chemische samenstelling en lading. De kop- en staartgroepen kunnen in een levende cel aan elkaar gekoppeld worden (waarvoor energie nodig is), zodat lange lineaire ketens ontstaan, de eiwitten.

Bijna ieder eiwit is ruimtelijk opgevouwen tot een goed gedefinieerde 3D-structuur. Deze 3D-structuur komt tot stand door zeer specifieke interacties van de zijketens van de aminozuren. Voor diverse eiwitten is deze 3D-structuur experimenteel bepaald met behulp van de technieken röntgendiffractie (van eiwitkristallen) of kernspinresonantie, NMR (van eiwitten in oplossing). Deze 3D-structuren kunnen met een computer op allerlei manieren worden weergegeven, waardoor prachtige plaatjes ontstaan, soms ook "molecular art" genoemd.

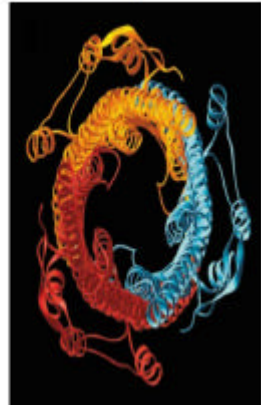
Talrijke eiwitten vervullen een specifieke functie in een organisme. Ik noem u eerst voorbeelden van "nuttige" eiwitten:

- bloedeiwitten, zoals hemoglobine, dat zuurstof bindt in de longen en vervoert naar alle delen van het lichaam waar deze overgedragen wordt aan een ander eiwit, myoglobine, in de spieren;
- antistoffen (immuunglobulinen) in bloed, die bescherming bieden tegen lichaamsvreemde stoffen door ze te binden en af te voeren;



- eiwitten van de oog lens, zogenaamde crystallines (een geliefd onderwerp van onderzoek bij de afdeling Biochemie van KUN), die door concentratiegradiënten zorgen voor lichtbreking en focusering op het netvlies;
- structuureiwitten, die structuur en sterkte geven aan bijvoorbeeld spinrag, zijde, haar, nagels, veren, huid, wol, spieren, en flagellen van bacteriën;
- transporteiwitten, die zorgen voor vervoer van ionen, suikers, aminozuren, mineralen, enz. vanuit een cel naar buiten, of omgekeerd;
- receptoreiwitten, die signalen opvangen en doorgeven;
- regulatie-eiwitten, die alle biochemische reacties in cellen coördineren en regelen.

Transporteiwit

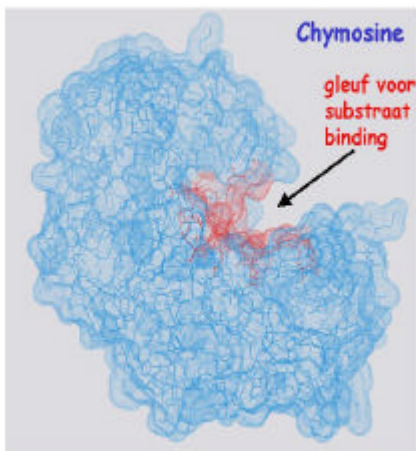


Er zijn ook veel voorbeelden van "slechte" eiwitten, zoals:

- BSE eiwitten, zogenaamde prionen, die van structuur veranderen als ze kwaadaardig worden;
- viruseiwitten, zoals HIV-protease en mond- en klauwzeervirus;
- toxine-eiwitten, zoals van slangen, spinnen en bacteriën.

Een heel belangrijke klasse eiwitten zijn de *enzymen*, die biochemische reacties versnellen. Binnen cellen zijn talloze enzymen actief die allerlei afbraak- en biosynthesereacties katalyseren. Gezuiverde enzymen worden ook al toegepast in industriële processen, zoals:

- in wasmiddelen: eiwit- en vetafbraak door proteases en lipases;
- in voedsel: melkstremming door chymosine bij kaasbereiding; verwerking van vruchten tot sap door pectinases, cellulases, amylases;
- voor bacteriedodend: celwandafbraak door lysozym.



Met een stereobril kunnen eiwitten op een computerscherm ook in 3D worden gezien. Sinds enkele jaren is het zelfs mogelijk om eiwitten in een virtuele omgeving (virtual environment) te bekijken, in de zogenaamde CAVE (Computer

Assisted Virtual Environment). Een CAVE is een kamer van 3 bij 3 meter, waarin een 3D-beeld geprojecteerd wordt (1). De onderzoeker staat hier middenin, met een speciale 3D-bril, en kan zo een eiwitmolecuul van alle kanten bekijken, en er ook binnenin kijken. Zo'n CAVE staat bij het Amsterdams rekencentrum SARA, en binnenkort komt er een alternatieve versie, die meer geschikt zal zijn voor onderwijs, in Groningen bij het nieuwe centrum voor High Performance Computing and Visualization (HPCV).

Variaties in eiwitstructuur en eigenschappen

Hoe komen eiwitten aan zo veel verschillende eigenschappen? Het geheim ligt in hun unieke 3D-structuur en in de grenzeloze variëteit daarin. Eiwitten bestaan uit ketens van zo'n 50-1000 van deze aminozuren. Het aantal mogelijke varianten is theoretisch 20^n , waarbij n het aantal bouwstenen, of aminozuren, is ($20^2 = 400$, $20^3 = 8000$, $20^4 = 160.000$, $20^5 = 3,2$ miljoen, $20^6 = 64$ miljoen, $20^7 = 1,28$ miljard, enz.). Hierdoor is een haast oneindig grote variëteit mogelijk. Deze koppeling van aminozuren is geen willekeurig proces. We zullen straks zien dat aminozuurvolgordes van eiwitten zeer nauwkeurig zijn vastgelegd in het erfelijke materiaal DNA, in een universele code die geldt voor alle levende organismen. Deze exact gedefinieerde volgorde en variëteit van de aminozuren in verschillende eiwitten bepalen de eigenschappen, of functies, van deze eiwitten.

Eiwitten kunnen worden ingedeeld in families. Het blijkt dat twee eiwitten die meer dan 25% identiteit vertonen in aminozuurvolgorde, ruimtelijk globaal op dezelfde manier zijn opgevouwen, en een zelfde of vergelijkbare functie uitoefenen. Hoe hoger het percentage identiteit in aminozuurvolgorde, des te groter de overeenkomst in 3D-structuur en functie. Straks zullen we zien dat deze constatering van cruciaal belang is voor het lezen van genetische kaarten. Als uit de aminozuurvolgorde blijkt dat een eiwit behoort tot een bepaalde familie, is het zelfs mogelijk om de 3D-structuur globaal te voorspellen, mits er van tenminste één familielid de 3D-structuur experimenteel is bepaald. Deze methode, die *homologie-modelling* heet, heb ik bij NIZO food research veel toegepast voor voorspelling van de structuur en katalytische eigenschappen van enzymen en andere eiwitten van melkzuurbacteriën. Helaas ontbreekt me nu de tijd om hier dieper op in te gaan, en ik verwijs u naar de groep van Prof. Gert Vriend van CMBI, die bestaat uit wereldexperts op dit gebied.

Dit waren voorbeelden van afzonderlijke eiwitten, maar eiwitten werken niet in afzondering. Er zijn duizenden verschillende eiwitten in levende cellen, en van ieder van die eiwitten zijn er duizenden kopieën in een cel. Ze worden voortdurend aangemaakt en na enige tijd weer afgebroken, recycling dus. Ze zijn onderdeel van uitgebreide netwerken, waarbij ze interacties aangaan met elkaar en met vele andere cellulaire componenten. Hierdoor ontstaat een levende, dynamische cel. De uitdaging is nu ten eerste om alle eiwitten die in een cel voorkomen (of beter gezegd *kunnen* voorkomen) en hun netwerken en interacties in kaart te brengen, en ten tweede om deze kaarten te lezen om er inzicht in te krijgen hoe een totale cel werkt, hoe alles in elkaar grijpt.

We still have a lot to learn about living organisms on earth.
(David Attenborough)

KAARTLEZEN

Zo kom ik bij het tweede woord van mijn titel, *kaartlezen*. Kaartlezen is voor mij altijd een sportieve passie geweest, en kaartlezen is nu ook in mijn wetenschappelijk onderzoek een belangrijke rol gaan spelen.

In mijn tijd als postdoc in Australië maakte ik kennis met de sport *orienteering* (oriënteringsloop in Nederland) (2), een combinatie van hardlopen, kaart- en kompaslezen, en ik was gelijk verkocht. In deze sport is het de bedoeling om een aantal controleposten te vinden, in de juiste volgorde, met als hulpmiddelen een kompas en een zeer gedetailleerde topografische kaart (schaal van 1:10.000 tot 1:50.000). Je doet dit om de beurt, en loopt meestal alleen, zeker in de wedstrijdsport waarbij het de bedoeling is om zo snel mogelijk rond te komen.

Dit speelt zich doorgaans af in prachtige natuurgebieden. In Australië betekent dit de “out-back”, wildernisgebieden met steile heuvels, veel rotsen en weinig paden. De *orienteering*-kaarten worden door de clubleden zelf gemaakt, en moeten zo nauwkeurig en gedetailleerd mogelijk zijn. Dus niet alleen paden, rivieren, meertjes en heuvels, maar ook hekken, soorten vegetatie, grote rotsblokken, bijzondere bomen, termietenheuvels, enz. De oriënteringsloopkaarten zijn bovendien 3-dimensionaal omdat heuvels en dalen met hoogtelijnen zijn aangegeven. Hoe nauwkeuriger de kaart, hoe gemakkelijker het is om je exacte positie te weten: als het ware kaartlezen “op de vierkante meter”.

Het beste leer je kaartlezen door te leren kaarten zelf te tekenen. De hoofdlijnen van een kaart (paden, wegen, rivieren) zijn gemakkelijk genoeg te tekenen, maar het nauwkeurig invullen van de details is een monnikenwerk, en kost vele uren veldwerk. Het lezen van Nederlandse kaarten is gemakkelijk, omdat onze natuurgebieden zoveel hoofdroutes, paden en ordening kennen. Door het ontbreken van hoogteverschillen is het zich oriënteren in Nederland een fluitje-van-een-cent. Maar in het buitenland zien de kaarten er vaak heel anders uit, met weinig paden of andere hoofdlijnen, veel heuvels en veel details over rotsen, kuilen enz. Het gevaar is dat je door de vele details eerder het overzicht en je richting verliest. Je ziet als het ware door de bomen het bos niet meer.

Orienteering is een ultieme combinatie van fysieke en mentale inspanning – een soort hardlopen en schaken tegelijk. Als beginner sprint je het bos in om binnen enkele minuten te constateren dat je op weg naar de eerste controlepost al verdwaald bent. Vervolgens leer je om voorzichtig van start te gaan, feeling met de kaart te krijgen, om daarna te versnellen als je de eerste posten gevonden hebt. Later leer je om tegen het einde van de loop, als je fysiek uitgeput raakt, de mentale concentratie vast te houden zodat je niet in het zicht van de haven strandt. Het is cruciaal om contact te houden met de kaart, ten allen tijde te weten:

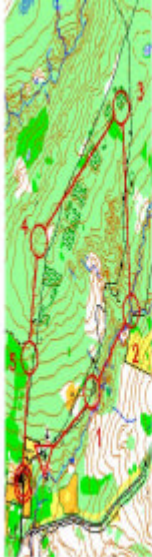
“Waar ben ik, waar kom ik vandaan, en waar wil ik naar toe”

Deze lessen en ervaringen komen ook uitstekend van pas in het dagelijks leven en in wetenschappelijk onderzoek.

Door *orienteering* leer je ook andere vaardigheden zoals:

- een plan van aanpak maken (routeplan, tempo);
- hulpmiddelen optimaal inzetten (kaart, kompas, herkenningspunten in het terrein);
- de omgeving goed analyseren (paden, heuvels, obstakels in het terrein);

- vele acties tegelijkertijd uitvoeren (hardlopen, kaartlezen, kompaslezen, omgeving analyseren, posten stempelen);
- snel beslissingen nemen onder tijdsdruk (route keuzes maken);
- snel trouble-shooten (verdwaald, wat nu: plan bijstellen of noodplan uitvoeren?);
- grote lijnen van details onderscheiden;
- overzicht houden;
- krachten goed verdelen (combinatie van fysieke en mentale vermogens);
- geheugen op scherp zetten (routeonderdelen uit het hoofd leren);
- na afloop: van je fouten leren (vergelijken van je gelopen route met die van je tegenstanders).



Als je dit enigszins onder de knie hebt kun je het ook verzwaren, bijvoorbeeld door het in teamverband te doen, in meerdere etappes, over meerdere dagen. In de vorm van *survival*/krijg je ook zwaardere fysieke en mentale opdrachten uit te voeren. Dan gaat teamwork een cruciale rol spelen: samen plannen en problemen oplossen, taken verdelen, wie is de zwakste schakel, fysiek en mentaal. Collega's van NIZO food research zullen zich onze deelname aan de Ede's Best Races herinneren van de beginjaren '90. Deze driedaagse survivals waren echt uitdagingen! We leerden teamwork, we leerden onder stress problemen oplossen, we leerden elkaars sterke en zwakke kanten kennen, we leerden onze eigen individuele grenzen kennen en verleggen. Maar *kaartlezen* was onze sterkte, ons geheim wapen.

Dit uitstapje naar landkaarten, plattegronden, topografische kaarten en *orienteering*-kaarten brengt mij nu naar biologische kaarten: genetische kaarten, metabolische kaarten en cellulaire kaarten.

GENOMEN

Genomen zijn geen tuinkabouters, dus ik hoop dat jullie je niet "genomen" voelen. Nee, genomen is het meervoud van "genoom", en dat is het totaal aan erfelijke (of genetische) informatie van een organisme vastgelegd in zijn DNA, of in zijn genen, zoals we dat vaak horen zeggen. Bestudering van afzonderlijke genen of groepen genen noemen we *genetica*, en bestudering van alle genen van een organisme tegelijk noemen we *genomica*, of *genomics*.

Levende cellen groeien, delen, communiceren met andere cellen en reageren op de buitenwereld. Hoe ze dat doen, wordt bepaald door de eigenschappen die opgeslagen liggen in hun erfelijk materiaal, in hun DNA. Dat DNA is bij hogere organismen verdeeld over meerdere *chromosomen*. Zo heeft bijvoorbeeld de mens 23 paar chromosomen, rijst heeft er 12, en gist 6. Bacteriën hebben doorgaans maar één chromosoom, dat circulair is.

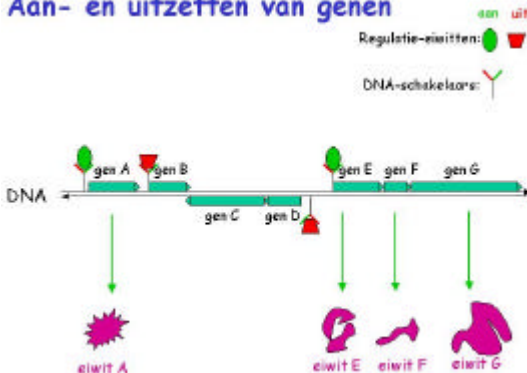
Het lezen van de genomkaart

De erfelijke informatie vastgelegd in DNA is in een code geschreven die eerst vertaald moet worden. Er zijn maar vier letters, namelijk A, C, G en T, de afkortingen voor de basen of nucleotiden adenine (A), cytosine (C), guanine (G) en thymine (T). Deze letters zijn niet op papier geschreven, maar ze zijn gekoppeld aan een drager, namelijk lange suiker-fosfaatketens, waardoor heel lange DNA-moleculen ontstaan. De lengte van deze keten kan zich uitstrekken van enkele honderdduizenden tot miljarden basen. Een chromosoom bestaat uit twee van deze lange ketens die in spiraalvorm om elkaar heen draaien, in een zogenaamde dubbele helix. Hierbij vormen zich de basenparen A-T en G-C, de strengen zijn dus exact complementair.

Je leest de letters in groepjes van drie, de zogenaamde codons, te vergelijken met korte woorden. Deze codons "vertalen" zich in de 20 verschillende aminozuren. Een lange reeks van deze codons, te vergelijken met een zin, codeert voor een eiwit. Deze stukken DNA die coderen voor eiwitten noemen we *genen*. Speciale start- en stop-codons geven aan waar genen beginnen en waar ze eindigen. De basenvolgorde in het DNA wordt op die manier vertaald in de aminozuurvolgorden van vele duizenden, verschillende eiwitten. DNA is dus als het ware een catalogus voor alle eiwitten die een cel zou kunnen maken.

Het is ook een catalogus voor de stukken DNA die tussen de genen liggen. Deze tussenliggende stukken zijn vaak verantwoordelijk voor de regulatie van de activiteit van de genen, m.a.w. hierin liggen de schakelaars die ervoor zorgen dat de genen aan- of uitstaan, dus of de gecodeerde eiwitten wel of niet gemaakt worden. Deze schakelaars worden aan- of uitgezet door speciale regulatie-eiwitten. Het DNA is dus naast een catalogus van eiwitten ook een handleiding voor de eiwittensynthese en cellulaire processen. Daarmee is een genomkaart een eerste stap op weg naar een volledig beeld van alle moleculaire activiteiten in een cel.

Aan- en uitzetten van genen



De DNA-code:
A, G, C, T



De cruciale stappen in het lezen van de genomkaart zijn:

- het bepalen van de volledige volgorde van alle basen in het DNA;
- het vinden van de genen en het vertalen van de DNA-volgorde in eiwitten;
- het vaststellen of voorspellen van de functies van die eiwitten;
- het vinden van de regulatie-elementen.

Genoomsequencing

Enkele decennia lang was de eerste stap, het bepalen van de DNA-volgorde, in vakjargon *DNA-sequencing* of *genoomsequencing*, erg tijdrovend en een belangrijke bottleneck. Een promovendus kon vier jaar besteden aan het bestuderen van slechts enkele genen en de gecodeerde eiwitten. Zo'n vijf jaar geleden kwam er een grote doorbraak toen er automatische DNA-sequencers op de markt kwamen die het mogelijk maakten om vele malen sneller de basenvolgorde in DNA te bepalen, de zogenaamde *high-throughput DNA-sequencing*. Nu werd het mogelijk de volledige DNA-volgorde van een genoom te bepalen en om alle genen van een organisme te vinden.

In 1995 werd de eerste volledige DNA-volgorde van een klein genoom gepubliceerd, iets minder dan 2 miljoen basenparen lang, dat van de influenzabacterie *Haemophilus influenzae* (3), door Craig Venter en medewerkers van TIGR, The Institute of Genome Research (4) en later zijn bedrijf Celera Genomics in de VS. Vanuit de hele wereld kwamen daar in de afgelopen vijf jaar meer dan 50 genoomvolgorden bij, voornamelijk van bacteriën die ziektes verwekken zoals tuberculose, cholera, dysenterie, ontstekingen van hersenvlies of longen, en infecties van wonden, luchtwegen, urinewegen en darmen (zie Tabel 1) (5). Verschillende andere genoomvolgordes zijn bijna klaar voor publicatie. Veel meer zijn al klaar, maar blijven geheim in de databanken van grote farmaceutische bedrijven.

Tabel 1. Genoomsequencing-projecten van ziekteverwekkende bacteriën.

Organisme	Ziekte	Publicatiejaar	Genoomgrootte (miljoen baseparen)
<i>Haemophilus influenzae</i>	influenza	1995	1,8
<i>Mycoplasma pneumoniae</i>	longontsteking	1996	0,8
<i>Escherichia coli</i> K-12	(laboratoriummodelstam)	1997	4,6
<i>Chlamydia trachomatis</i>	urinewegeninfecties	1998	1,7
<i>Mycobacterium tuberculosis</i>	tuberculose	1998	4,4
<i>Treponema pallidum</i>	syfilis	1998	1,1
<i>Chlamydia pneumoniae</i>	luchtwegeninfecties	1999	1,2
<i>Helicobacter pylori</i>	maagzweer	1999	1,6
<i>Neisseria meningitidis</i>	hersenvliesontsteking	2000	2,2
<i>Pseudomonas aeruginosa</i>	luchtwegeninfectie	2000	6,2
<i>Ureaplasma urealyticum</i>	urinewegeninfectie	2000	0,8
<i>Vibrio cholerae</i>	cholera, diarree	2000	2,5
<i>Escherichia coli</i> O157:H7	urinewegen- en darminfecties	2001	4,1
<i>Mycobacterium leprae</i>	lepra	2001	3,3
<i>Enterococcus faecalis</i>	wond-, luchtwegen- of urinewegen- infectie, hersenvliesontsteking	2001?	3,2
<i>Neisseria gonorrhoeae</i>	urinewegeninfectie	2001?	2,2
<i>Salmonella typhi</i>	darminfectie	2001?	4,5
<i>Salmonella typhimurium</i>	darminfectie	2001?	4,5
<i>Staphylococcus aureus</i>	wondinfectie	2001?	2,8
<i>Streptococcus pneumoniae</i>	hersenvliesontsteking, sinusitis	2001?	2,2
<i>Streptococcus pyogenes</i>	keelinfectie, diarree	2001?	1,9

Tabel 2. Genoomsequencing-projecten van hogere organismen.

Organisme		Aantal chromosomen	Genoomgrootte (miljoen baseparen)	Aantal genen	Publicatie- jaar
bacterie	<i>Haemophilus influenzae</i>	1	2	1.850	1995
gist	<i>Saccharomyces cerevisiae</i>	16	12	6.300	1997
rondworm	<i>Caenorhabditis elegans</i>	6	97	19.000	1998
fruitvlieg	<i>Drosophila melanogaster</i>	6	137	14.000	2000
zandraket (plant)	<i>Arabidopsis thaliana</i>	5	120	25.000	2000
rijst	<i>Oryza sativa</i>	12	430	ca. 50.000	?
mens	<i>Homo sapiens</i>	23	3.500	ca. 30.000	2001

Genomen van hogere organismen zijn veel groter dan die van bacteriën, waardoor het langer duurt om hun DNA-volgorde te bepalen. Bovendien is het lezen van hun genoomkaarten moeilijker omdat de genen in fragmenten opgesplitst zijn (in zogenaamde exons, onderbroken door introns) die eerst bij elkaar gezocht moeten worden. De grootte van het genoom blijkt niet evenredig te zijn met het aantal genen: hoe groter het genoom hoe geringer de dekking met genen (zie Tabel 2). In bacteriën codeert ca. 90% van het DNA voor genen, bij de mens is dat nog maar 1%. De rest is grotendeels van nog onbekende functie, en wordt *junk-DNA* genoemd. Toch zijn er in de

afgelopen twee jaar al kladversies van de genomen van de mens, een vlieg, een worm en enkele planten gepubliceerd, en daar komen binnenkort rijst, een muis, een rat en een vis bij.

De media besteden de laatste maanden veel aandacht aan genoomsequencing, met name van het menselijk genoom. Science Magazine, het belangrijkste wetenschappelijke vakblad van de VS, riep "genome sequencing" uit tot de belangrijkste wetenschappelijke doorbraak van 2000 (6). Time Magazine riep George Bush uit tot "Person of the Year", maar de runner-up was "Gene Mapper" Craig Venter, die verschillende doorbraken realiseerde voor het in kaart brengen van het menselijk genoom.



Breakthrough of year 2000



Having the (human) genome in hand will almost certainly be seen as one of the crowning achievements of the new century, no matter what else happens in the next 100 years. The genome – or more precisely, the individual genes it contains – spell out the instructions for constructing the protein building blocks of every cell in every tissue of the body. This so-called book of life will inevitably reveal secrets of both health and disease, promising new treatments for virtually every malady that afflicts us.
(TIME Magazine, January 1, 2001)

BIO-INFORMATICA

Het gereedschap dat we gebruiken om genoomkaarten te interpreteren, heet *bio-informatica*. Genoomsequencing-projecten leveren gigantische hoeveelheden data op en er komt in hoog tempo nieuwe informatie bij. Om deze vloedgolf aan biologische gegevens te verwerken heb je de gereedschappen, de tools van bio-informatica nodig. Bio-informatica is een relatief nieuwe discipline die zich beweegt op het grensvlak van biologie, computerwetenschappen en toegepaste wiskunde. Bio-informatica houdt zich bezig met de ontwikkeling en toepassing van methoden uit de informatica voor de verzameling en opslag van biologische gegevens, en vervolgens de analyse, interpretatie, het beschikbaar maken en toepassen. Waar voorheen het onderzoek beperkt bleef tot “beschrijven en verklaren”, zal met het genoomonderzoek de doorbraak moeten komen naar “voorspellende kennis”. Ondersteuning vanuit de bio-informatica is hierbij essentieel om de toenemende stroom aan biologische data te integreren in databanken, te verwerken tot modellen en te transformeren naar kennis.

Bio-informatica maakt gebruik van computers, specifieke software, databanken, het internet en zoekmachines. Dit noemen we *in-silico*-analyse, in tegenstelling tot experimenten in reageerbuizen (*in vitro*) of in levende organismen (*in vivo*).

Het Center for Molecular and Biomolecular Informatics (CMBI, voorheen CAOS/CAMM) onder de huidige leiding van Prof. Dr. Gert Vriend is al vele jaren het centrale instituut in Nederland voor bio-informatica-expertise. De Bio-informatica, Databases en Data Mining faciliteit van CMBI onder leiding van Dr. Jack Leunissen geniet grote bekendheid. Het instituut beschikt over krachtige computers om honderden databanken en software tools (diverse in licentie) te beheren, aan elkaar te koppelen en te implementeren voor ondersteuning van wetenschappelijk en toegepast onderzoek (7).

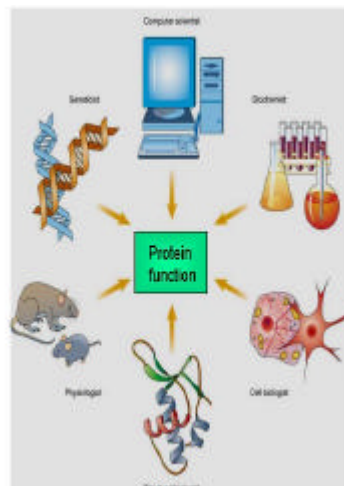
Databanken

Al deze biologische gegevens worden opgeslagen in grote centrale databanken die via internet beschikbaar zijn. Er zijn momenteel meer dan 500 biologische databanken die via het web toegankelijk zijn (8, 9). Vele hiervan bevatten algemene gegevens zoals DNA-volgorden, eiwitaminozuurvolgorden en eiwitstructuren. Dit zijn databanken die dagelijks hun gegevens uitwisselen en vrij toegankelijk zijn voor alle onderzoekers. De EMBL-databank is Europa's primaire bron van DNA gegevens. De SWISS-PROT-databank (10) richt zich op eiwitten en bevat gedetailleerde informatie over annotatie (= functietoekenning). Inmiddels bevat deze databank informatie van meer dan 100.000 verschillende eiwitten. TREMBL is een eiwitdatabank met meer dan 250.000 eiwitsequenties, verkregen door vertaling van de DNA-volgorden in de EMBL-databank. Er zijn ook databanken die specialistische informatie bevatten, bijvoorbeeld over afzonderlijke genomen, families van eiwitten, moleculaire functies, metabolische routes, gen-expressie of genregulatiefactoren (zie Tabel 3).

Voorspelling van eiwitfunctie

Terug naar het lezen van de genoomkaart. Als je de volledige DNA-volgorde bepaald hebt van een genoom, en op basis van start- en stopcodons voorspeld hebt waar potentiële genen liggen, dan is de volgende stap om deze genen te vertalen in de aminozuurvolgorden van eiwitten. Dat is eenvoudig, want de code is bekend. Dit is het begin van je catalogus van eiwitten. Dan begint pas het moeilijke werk, waarbij bio-informatica een cruciale rol gaat spelen. Je wilt nu aanwijzingen zien te krijgen over de functies die deze eiwitten zouden kunnen spelen in de cel. Dit noemen we *annotatie*, of functiebeschrijving. Dat kun je experimenteel op allerlei verschillende manieren doen, maar je kunt eenvoudig beginnen met *functievoorspelling*.

De snelste manier om aanwijzingen te krijgen is de basenvolgorde van je onbekende DNA-basenvolgorden en eiwit-aminozuurvolgorden aanwezig in de beschikbare databanken. Deze methode wordt ook wel *BLASTen* genoemd, naar de naam van het populaire softwarepakket (11). Het gaat om honderdduizenden tot miljoenen vergelijkingen, maar is met de PC in seconden gebeurd! Op deze manier vind je snel verwante eiwitten, meestal uit andere organismen, waarvan de cellulaire functie en soms zelfs de 3D-structuur al bekend is. Deze verwantschap is eigenlijk een direct gevolg van evolutie. In de loop van de evolutie verandert de basen-volgorde in DNA heel geleidelijk door natuurlijke mutaties. Door deze verandering van de codons, veranderen de aminozuurvolgorden ook geleidelijk, maar in de praktijk blijkt dit niet gauw tot verandering van de functie van zo'n eiwit te leiden.



Tabel 3. Voorbeelden van belangrijke biologische databanken.

Databanken	Inhoud	Website
<i>Algemeen</i>		
EMBL	DNA-volgorden	www.ebi.ac.uk/embl/
SWISS-PROT	eiwitvolgorden	www.expasy.ch/sprot/
TREMBL	eiwitvolgorden	www.ebi.ac.uk/trembl
PDB	3D-structuren van eiwitten	www.rcsb.org/pdb/
<i>Genomen</i>		
ECDC	<i>E.coli</i>	susi.bio.uni-giessen.de/ecdc/
SGD	gist	genome-www.stanford.edu/
Flybase	fruitvlieg	fly.ebi.ac.uk:7081/
CMR	microbiële genomen	www.tigr.org/tigr-scripts/CMR2/
<i>Moleculaire functies</i>		
EMP	enzymfunctie	wit.mcs.anl.gov/WIT2/
Transfac	regulatiefactoren	transfac.gbf-braunschweig.de/TRANSFAC/
DIP	eiwitinteracties	dip.doe-mbi.ucla.edu
<i>Eiwitfamilies</i>		
Pfam	eiwitfamilies	www.sanger.ac.uk/software/Pfam
COG	orthologe eiwitten	www.ncbi.nlm.nih.gov/COG
ProDom	eiwitdomeinen	protein.toulouse.inra.fr
CATH	3D-structuurfamilies	www.biochem.ucl.ac.uk/bsm/dhs
<i>Eiwitmotieven (vingerafdrukken)</i>		
ProSite	eiwitfunctiemotieven	www.expasy.ch/sprot/prosite.html
Prints	eiwitfunctiemotieven	www.biochem.ucl.ac.uk/bsm/dbbrowser/PRINTS
Blocks	eiwitfunctiemotieven	www.blocks.fhcrc.org
<i>Cellulaire processen</i>		
KEGG	metabolische routes	www.genome.ad.jp/kegg/
WIT	metabolische routes, transport	wit.mcs.anl.gov/WIT2/
Expasy	metabolische routes	www.expasy.ch/cgi-bin/search-biochem-index
EcoCyc/MetaCyc	metabolische routes (<i>E.coli</i>)	
aMAZE	metabolische routes, regulatie, transport	www.ebi.ac.uk/research/pfbbp/
<i>Integratie databanken</i>		
SRS	linked databanken	srs.ebi.ac.uk
Entrez	linked databanken	www.ncbi.nlm.gov/Entrez/
Interpro	eiwitdatabanken	www.ebi.ac.uk/interpro/

Eerder in mijn verhaal vertelde ik dat eiwitten zelfs al tot dezelfde familie behoren als hun aminozuurvolgorden voor meer dan 25% identiek zijn. Hoe hoger het percentage identiteit in aminozuurvolgorde, des te groter de overeenkomst in 3D-structuur en de kans dat ze dezelfde

functie hebben. Twee organismen die evolutionair dicht bij elkaar staan, hebben genen die maar weinig verschillen in basenvolgorde, en eiwitten die maar weinig van elkaar verschillen in aminozuurvolgorde. In de praktijk blijkt dat men op basis van deze evolutionaire verwantschappen al een functie kan voorspellen voor 50% van de gecodeerde eiwitten, en soms meer.

Er zijn ook andere kenmerken van eiwitten die we kunnen gebruiken voor functievoorspelling. Zo kennen we korte geconserveerde aminozuurvolgorden – die we *motieven*, *patronen* of *fingerprints* noemen – waarvan bekend is dat ze essentieel zijn voor een bepaalde activiteit of functie of locatie van een eiwit in de cel. Dit kunnen bijvoorbeeld katalytische residuen van enzymen zijn, of residuen betrokken bij de binding van substraten, metaalionen, co-factoren, enz. Ook kun je voorspellen of een eiwit in het cytoplasma, dus in de cel, zal blijven, in het celmembraan terecht zal komen, of juist uit de cel getransporteerd zal worden.

Om zulke voorspellingen te kunnen doen is er in de bio-informatica een heel scala van databanken en softwaretools ontwikkeld (ref. 8, 12; Tabel 3). Bij gebruik van al deze methodes is het natuurlijk essentieel om de mogelijkheden en beperkingen van functie- of structuurvoorspelling goed af te wegen (13).

Uit de gepubliceerde genomen van zo'n 30 bacteriën is gebleken dat je van gemiddeld 60% van de eiwitten een functie kunt voorspellen. Van de rest komt 15% voor in meerdere genomen, de zogenaamde *conserved hypothetical proteins*, maar is de functie nog nergens bekend, en gemiddeld 25% van de eiwitten zijn uniek voor iedere bacterie (14). Deze voorspellingen vormen de basis voor experimentele verificatie. Voor de eerste groep eiwitten zal de voorspelde functie experimenteel bevestigd moeten worden, en van de tweede groep zal dit de enige manier zijn om achter de functies te komen.

Reconstructie van de metabolische kaart en andere cellulaire processen

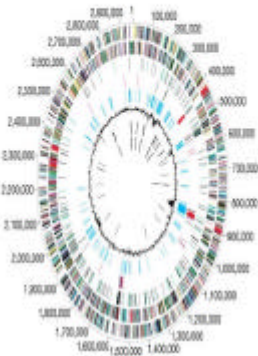
De volgende stap in het reconstrueren van cellulaire processen uit de genoomgegevens is het classificeren van de eiwitten met voorspelde functies, dus indelen in de biologische processen waar ze bij betrokken zijn. In andere woorden, je wilt de gevonden eiwitcatalogus indelen in rubrieken. In de literatuur zijn al veel verschillende classificatiesystemen beschreven en vergeleken (15), maar helaas heeft er nog geen standaardisatie plaatsgevonden. Een voorbeeld van rubrieken is:

1. Aminozuurbiosynthese
2. Biosynthese van co-factoren, prosthethische groepen
3. Celenvelop
4. Cellulaire processen
5. Centraal intermediair metabolisme
6. Energiemetabolisme
7. Vetzuur- en fosfolipidemetabolisme
8. Purines, pyrimidines, nucleosides en nucleotides
9. Regulatie
10. Replicatie
11. Transcriptie
12. Translatie
13. Transport
14. Andere categorieën
15. Hypothetische eiwitten

Dit systeem heeft maar één algemeen niveau, maar verdieping met meer gedetailleerde niveaus is mogelijk. Als je deze rubrieken nu met verschillende kleuren intekent op de genomkaart, begin je al patronen van gen-organisatie te zien. Je ontdekt clusters van genen die dezelfde functie hebben, die zogenaamd functioneel gekoppeld zijn. Deze kunnen bijvoorbeeld coderen voor alle enzymen van een biosynthese-route, of de subeenheden zijn van een enzymcomplex. Dikwijls blijven zulke clusters in de loop van de evolutie intact, en zijn ze terug te vinden in genomen van verwante bacteriën. Een volledig bacterieel genoom wordt meestal weergegeven als een cirkel met gekleurde streepjes voor de verschillende genen.

Genoomorganisatie

Vibrio cholerae (chromosoom I)

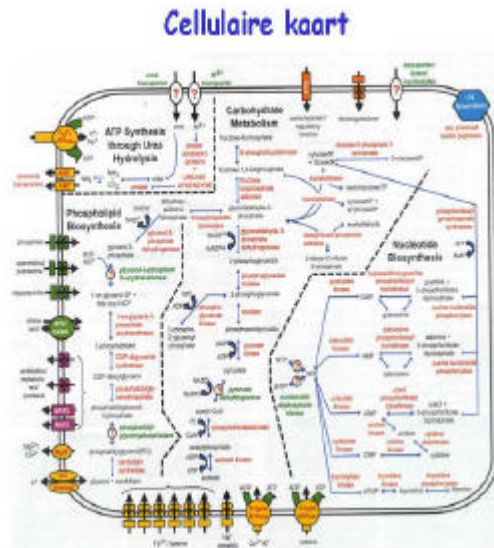


Met zo'n indeling in functionele klassen kunnen vervolgens metabolische routes – de afbraak- en biosynthese-routes – in kaart gebracht worden en (delen van) het metabolische netwerk van de cel, bestaande uit de routes, de enzymen en de metabolische producten (substraten, activatoren, inhibitoren, enz) (16). Deze informatie kan ook opgeslagen worden in metabolische databanken (17), waarvan EcoCyc DB voor *E. coli* momenteel de meest complete set bevat (18). Andere metabolische databanken zoals KEGG (19) en WIT (20) bevatten algemene overzichten van metabolische routes, voornamelijk gebaseerd op alle bacteriële genomen waarvan de DNA-volgorde is bepaald.

De meest uitgebreide voorspelling van metabolisch potentieel uit volledige genoomdata tot nu toe is uitgevoerd voor de bacterie *Haemophilus influenzae* (21), waarbij 461 metabolische reacties met 367 intermediären en 84 producten, onderverdeeld in zes metabolische netwerken, in kaart zijn gebracht. Tenslotte kun je proberen om een totale cellulaire kaart te maken waarin de metabolische routes, de transportsystemen, de signaaloverdrachtsystemen en zelfs de genregulatiesystemen zijn opgenomen.

We moeten ons voortdurend realiseren wat de beperkingen zijn van deze voorspellingen en simulaties. Enerzijds blijven het natuurlijks voorspellingen van eiwitfuncties en hun samenhang, en anderzijds is het plaatje ook nog eens incompleet, omdat je voor zo'n 30-40% van de eiwitten geen functie kunt voorspellen. Bovendien levert het een statisch plaatje op van de cel. Je weet welke eiwitten gemaakt *kunnen* worden, maar het is nog niet dynamisch. Je weet nog niet welke genen aan- of uitstaan onder diverse externe omstandigheden, en welke metabolische routes actief zijn. Dit hangt grotendeels af van allerlei regulatie-mechanismen, de aan- en uitschakelaars als het ware, en uit ervaring weten we dat dit een zeer complex netwerk kan zijn.

Maar gelukkig zijn er de afgelopen jaren andere snelle, *high-throughput*-methoden ontwikkeld die experimentele ondersteuning kunnen bieden. *Transcriptomics* is een methode om met behulp van DNA-chips (of microarrays) de activiteit van grote aantallen genen (bijvoorbeeld van een heel genoom) tegelijkertijd vast te stellen. Dit gebeurt door het kwantificeren van RNA-moleculen, de afschriften van genen, die een tussenstap zijn op weg naar de synthese van eiwitten. Dit kun je onder verschillende experimentele omstandigheden uitvoeren, en zodoende een dynamisch beeld krijgen welke genen aan- en uitstaan.



Daarnaast kun je met *proteomics* een beeld krijgen van het totale eiwitpatroon dat onder verschillende omstandigheden in een cel aanwezig is. Deze 2D-elektroforesetechniek levert een heel direct beeld op welke eiwitten werkelijk aangemaakt worden en in welke relatieve hoeveelheden. Je moet natuurlijk wel de eiwitten in de spots op de gel kunnen identificeren en relateren aan de aanwezige genen in het genoom, maar daar zijn nu snelle, geavanceerde massaspectrometrische methoden voor ontwikkeld. Zowel *transcriptomics* als *proteomics* leveren ook gigantische hoeveelheden data op, die met bio-informaticatools verwerkt kunnen worden.

Met weer andere technieken kun je aantonen dat ieder eiwit interacties aangaat met 5 tot 50 andere eiwitten in een cel (22, 23). Dit kunnen interacties zijn in een eiwitcomplex, zoals het ribosoom, in dezelfde metabolische route, zoals in biosyntheseroutes, of in hetzelfde biologische proces, zoals regulatie en signaaloverdracht. Ook hier kun je spreken van het in kaart brengen van de eiwitinteractienetwerken in een cel. Al deze experimenten hebben ons geleerd dat we een dynamische cel moeten zien als sets van genen, RNA's en eiwitten die afhankelijk van de experimentele omstandigheden (het milieu), op een heel specifieke manier wat betreft ruimte, tijdstip en hoeveelheid, aanwezig zijn in de cel.

Tenslotte is een laatste ontwikkeling in de bio-informatica het integreren van verschillende soorten gegevens verkregen uit enerzijds *in-silico*-genoomanalyse en anderzijds experimenteel onderzoek (*transcriptomics* en *proteomics*) om metabolische modellen op te stellen voor de

simulatie van intracellulaire metabolische processen, om daarmee voorspellingen te doen over dynamisch gedrag van cellen. Deze aanpak moet leiden naar de *virtuele cel* die bestaat uit een softwarepakket waarmee genen, eiwitten en andere moleculen en hun individuele interacties beschreven kunnen worden. Een voorbeeld is de software E-CELL (24), waarmee een virtuele cel werd gesimuleerd met slechts 127 genen, voldoende voor transcriptie, translatie, energieproductie en fosfolipidesynthese, van *Mycoplasma genitalium*. Dit model is ook uiterst geschikt voor simulaties van voedselbacteriën met een klein genoom, zoals melkzuurbacteriën. Dergelijke modellen gebruiken als input ook biochemische en fysiologische data, ondermeer verkregen uit *transcriptomics* en *proteomics*. Zulke modellen zijn zeer waardevol voor voorspellingen hoe het metabolisme gestuurd kan worden, en vormen de basis voor *metabolic-engineering*-experimenten.

Hoeveel van zo'n genetische of metabolische kaart heb je nu eigenlijk nodig, hoeveel moet je in kaart brengen? Dat hangt natuurlijk af van je wetenschappelijke vraagstellingen. Als voorbeeld: Als je de kortste route van Ede naar Nijmegen wilt weten, heb je niet de hele topografische kaart van Nederland nodig. Maar als je wilt weten welke fietspaden er zijn in Nederland, dan heb je een complete gedetailleerde kaart nodig. Het is dus zaak om je wetenschappelijke vragen altijd goed te formuleren, en alleen die stukken van de genetische of metabolische kaart uit te zoeken die je nodig hebt.

Vergelijkend genomonderzoek

Onder vergelijkend genomonderzoek (*comparative genomics*) wordt verstaan de vergelijking van volledige genomsequenties (en daarvan afgeleide transcriptoom- en proteoomdata) van verschillende bacteriën. Hiermee kunnen overeenkomsten en verschillen geïdentificeerd worden, onder meer in biosyntheseroutes, transportsystemen, metabolische routes en regulatienetwerken. Door het voorhanden zijn van steeds meer complete genomsequenties zijn bacteriële genomen uitermate geschikt voor het toepassen van vergelijkend genomonderzoek.

Vergelijking van genomen kan inzicht geven in genfusies of overeenkomsten in clustering van genen, en kan erop wijzen dat deze genen gefuseerd zijn of tijdens hun evolutie omwille van functionele afhankelijkheid selectief bij elkaar gehouden zijn (25). Het kan ook een belangrijke aanvulling zijn voor functietoekenning van genen.

Vergelijkend genomonderzoek staat nog in het begin van zijn ontwikkeling en vereist veel ondersteuning vanuit de bio-informatica. Het software programma STRING (26) geeft snel een overzicht van geconserveerde genclusters in verschillende genomen, en is ontwikkeld binnen de groep van de Nederlander Martijn Huynen, één van 's werelds experts in vergelijkend genomonderzoek en genomevolutie (27, 28).

Vergelijkend genomonderzoek zal uitmonden in wetenschappelijk inzicht in de evolutie en diversiteit van soorten, stammen en genen. Inzicht in gemeenschappelijke omzettingen en routes is belangrijk voor de voorspelling van het gemeenschappelijk gedrag van de diverse soorten en stammen. Unieke genen of metabolische routes daarentegen geven soorten of stammen hun unieke positieve of negatieve functionele eigenschappen, en hierin ligt een grote uitdaging om deze genen te identificeren en te benutten of juist als targets te gebruiken voor bestrijding.

GENOOMONDERZOEK VAN INDUSTRIEEL-RELEVANTE BACTERIËN

Sinds enkele jaren worden ook de genomen van niet-ziekteverwekkende bacteriën ontrafeld (14). De keuze valt op nuttige, industrieel-relevante bacteriën (Tabel 4), zoals voor:

- bio-remediatie, het oplossen van milieuproblemen: recycling, afvalreiniging (water, grond, lucht);
- natuurlijke producten: metabolieten, insecticiden, antibiotica, stabiele biokatalysatoren (enzymen);
- bio-mining: winning en oxidatie van metalen, biogassen;
- landbouw: stikstofbinding, fotosynthese, bio-energie;
- humane en dierlijke voeding: spijsvertering, fermentatie, vergisting, smaak en textuur, gezondheid.

De Department of Energy (DOE) in de VS heeft een Joint Genome Institute Microbial Sequencing (29) opgericht en financiert al vanaf 1994 een Microbial Genome Project (30) met de nadruk op tientallen bacteriën die een rol spelen bij bijvoorbeeld energieontwikkeling, afbraak van chemische afvalstoffen, zuivering van afvalwater, katalyse van industriële processen, enz. Hun onlangs gestart Microbial Cell Project (31) heeft tot doel inzicht te ontwikkelen in de wijze waarop cellulaire componenten samenwerken om levende systemen te vormen.

Bacteriën en ons voedsel

Ik zal me voornamelijk bezig gaan houden met voedsel-relevante bacteriën, die uiteen vallen in de "bad bugs" en de "good bugs".

Tabel 4. Voorbeelden van genomesequencing van industrieel relevante bacteriën.

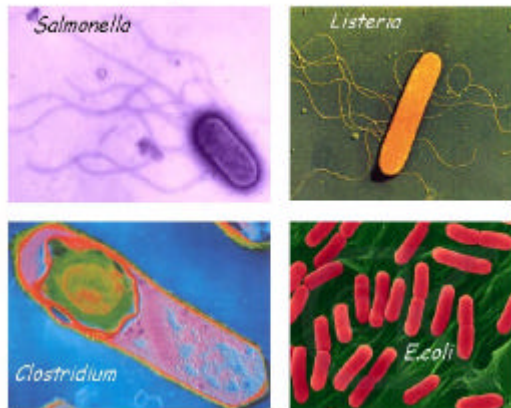
Micro-organisme	Toepassing	Gereed	Genoomgrootte (miljoen baseparen)
Bioremediatie			
<i>Deinococcus radiodurans</i>	verwerking radioactief afval	1999	3,2
<i>Pseudomonas putida</i>	afbraak polymeren, organisch afval	2000	6,2
Natuurlijke producten			
<i>Bacillus halodurans</i>	biokatalysatoren	2000	4,2
<i>Streptomyces coelicolor</i>	natuurlijke antibiotica	2000	8,0
<i>Photobacterium luminescens</i>	natuurlijke insecticiden	2000	5,5
Bio-mining, energie			
<i>Methanococcus jannaschii</i>	biogasproductie	1996	1,7
<i>Archeoglobus fulgidus</i>	sulfaatreductie	1997	2,2
<i>Thermotoga maritima</i>	afbraak plantpolymeren, energie	1999	1,9
Landbouw			
<i>Rhodospirillum rubrum</i>	fotosynthese	2000	3,7
<i>Clostridium thermocellum</i>	afbraak plantenpolysacchariden	2001?	?

Bad bugs: schadelijke voedselbacteriën

De "bad bugs", bacteriën die de voedselveiligheid en -kwaliteit in negatieve zin kunnen beïnvloeden, zijn de voedselpathogenen en bederfororganismen zoals *Escherichia coli*, *Campylobacter*, *Salmonella*, *Listeria*, *Bacillus cereus*, en *Clostridia*. Ze veroorzaken vooral infecties, ontstekingen, vergiftiging en diarree.

Slechte bacteriën maken vaak eiwitten waarmee ze zich kunnen hechten in darmen, lucht- of urinewegen, om vervolgens daar uit te groeien. Daarnaast kunnen ze toxines maken die de gastheercel binnendringen, of andere gunstige bacteriën afdoden. Deze "bad bugs" bezitten ook allerlei mechanismen en eiwitten voor overleving binnen of buiten de gastheer; we noemen dat resistentie tegen stressfactoren zoals zuurstof, extreme temperatuur, zuurgraad of zoutconcentratie, en natuurlijk ook antibioticumresistentie. Met genoomonderzoek probeert men de verantwoordelijke genen, eiwitten en mechanismen te identificeren, zodat deze als mikpunt kunnen dienen voor gerichte bestrijding.

Schadelijke voedselbacteriën



Tabel 5. Voorbeelden van genoomsequencing van schadelijke voedselbacteriën.

Bacterie	Bederf of ziekte	Publicatie- jaar	Genoomgrootte (miljoen baseparen)
Gram-positief			
<i>Listeria monocytogenes</i>	listeriose, hersenvliesontsteking	2001	2.9
<i>Bacillus cereus</i>	diarree	2001?	5.4
<i>Clostridium botulinum</i>	botulisme	2001?	4.0
<i>Enterococcus faecalis</i>	wond-, luchtwegen- en urinewegen- infectie, hersenvliesontsteking	2001?	3.2
<i>Clostridium perfringens</i>	darminfectie	2001?	?
Gram-negatief			
<i>Campylobacter jejuni</i>	koorts, hoofdpijn, diarree	2000	1.6
<i>Escherichia coli</i>	urinewegen- en darminfectie	2001	4.1
<i>Salmonella enteritidis</i>	voedselvergiftiging	2001?	4.5
<i>Salmonella typhimurium</i>	voedselvergiftiging	2001?	4.5
<i>Shigella flexneri</i>	diarree	2001?	4.7
<i>Yersinia enterocolitica</i>	darminfectie	2001?	4.4



Voor voedselveiligheid is het van eminent belang in een vroeg stadium pathogene en bederfbacteriën te identificeren, typeren, kwantificeren en bestrijden. Door de toenemende beschikbaarheid van genoomsequenties van deze micro-organismen zullen geautomatiseerde typeringsmethodes worden ontwikkeld. DNA-microarrays met specifieke merkers (probes) voor

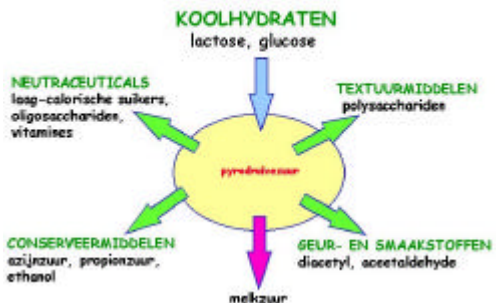
duizenden verschillende soorten of stammen zullen een snelle en betrouwbare identificatie van micro-organismen in levensmiddelen mogelijk maken. Genoomsequenties van enkele voedsel-pathogenen en bederfbacteriën zijn al beschikbaar, en vele andere naderen hun voltooiing in *Salmonella* en *Clostridium difficile*, de VS (*Campylobacter* en *Staphylococcus aureus*) en Frankrijk (*Listeria*) (zie Tabel 5).

In januari 2001 is het genoom van de bacterie *Escherichia coli* stam O157:H7 ontrafeld (32). Deze bacterie, die soms in rauwe en niet goed bereide hamburgers voorkomt, veroorzaakt in de VS jaarlijks vele doden. Het aantal ziektegevallen is 75 duizend per jaar in de VS en 20-30 in Nederland. Mensen die besmet zijn, krijgen bloederige diarree en lopen schade op aan de nieren. De infectie treft vooral kinderen, ouderen en mensen met een verminderde werking van het immuunsysteem. De bacterie begint steeds resistenter te worden tegen antibiotica. Met het in kaart brengen van het genoom is de bestrijding van deze bacterie een stap dichterbij gebracht. Het genoom van *E. coli* O157:H7 is vergeleken met de in laboratoria veel gebruikte onschadelijke *E. coli* stam K-12. De twee stammen blijken zo'n 3500 overeenkomstige genen te hebben, maar O157:H7 heeft daarnaast nog 1300 extra genen die niet voorkomen in K-12. Mogelijk kunnen die verschillen uitsluitsel geven waarom de ene stam gevaarlijk is en de andere onschuldig.

Prof. Wiel Hoekstra zei hierover in Bionieuws (33): "De gevonden DNA-volgorde legt in principe alle geheimen van de bacterie bloot, en dus ook de achillespezen. Het is, om in militaire termen te spreken, alsof je de plattegrond van een vijandelijke stad in handen hebt gekregen. Met enig speurwerk kun je dan als strateeg uitvlooien waar troepen moeten infiltreren of waar bommen moeten vallen. Zo kan ook een celbioloog aan de hand van de DNA-volgorde en met het nodige vernuft ontdekken met welke middelen specifieke en vitale functies van de ziekteverwekkende bacterie uitgeschakeld kunnen worden."

Op het gebied van voedselpathogenen is er specifieke kennis van gram-positieve bacteriën ontwikkeld bij met name Wageningen Universiteit, TNO-Voeding en NIZO food research; deze is nu deels gebundeld binnen het WCFS, het Wageningen Center for Food Sciences. In de

Melkzuurbacteriën: "Suikerstofwisseling"



toekomst zullen we ons concentreren op vergelijkend genomonderzoek van enkele relevante gram-positieve bederfbacteriën en pathogenen, zoals *Listeria*, bacilli en clostridia.

Good bugs: nuttige bacteriën

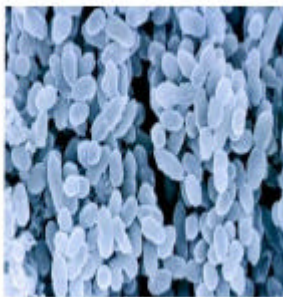
Bij de productie van gezonde voedingsmiddelen en -ingrediënten ligt een groeiende nadruk op de functionaliteit die micro-organismen bij fermentaties toevoegen aan het uitgangsmateriaal door de productie van conserverende verbindingen (zuren), smaak- en geurstoffen, textuurvormende polymeren zoals exopolysacchariden, en gezondheidsbevorderende componenten, die we met een mooi woord *neutraceuticals* noemen. Met name de zogenaamde "gram-positieve" bacteriën omvatten industriële organismen voor productie van voedingsmiddelen of ingrediënten, zoals enzymen en metabolieten, die alleen al in Nederland betrokken zijn bij een jaarlijkse omzet van meer dan 5 miljard gulden. Deze omvatten soorten van *Lactococcus*, *Lactobacillus*, *Streptococcus* en *Pediococcus*, melk-zuurbacteriën die gebruikt worden bij de bereiding van gefermenteerde zuivelproducten zoals kaas, karnemelk en yoghurt maar ook bij de productie van ingrediënten, en *Bacillus*-soorten die gebruikt worden voor de productie van industriële food- en non-food-enzymen en andere ingrediënten.

Daarnaast omvatten de gram-positieve bacteriën een grote groep organismen die na consumptie belangrijke effecten op de gezondheid kunnen hebben, met name in het maag-darmkanaal. Specifieke functionaliteit als hechting, overleving en productie van specifieke gezondheidsbevorderende verbindingen is essentieel voor het gebruik van deze micro-organismen als *probiotica*. Een scala aan positieve effecten wordt toegeschreven aan melk-zuurbacteriën – met name soorten van *Lactobacillus*, *Lactococcus* en *Bifidobacterium* – met een probiotische of andere gezondheidsbevorderende werking zoals het afleveren van therapeutische moleculen in het maag-darmsysteem (34).

In de zuivelindustrie wordt (gepasteuriseerde) melk op een gecontroleerde manier met behulp van speciaal toegevoegde bacteriën omgezet tot producten als yoghurt en kaas. Hierbij wordt de melksuiker lactose vergist en treedt de bekende verzuring op, waardoor de houdbaarheid van het product toeneemt. Het verzuurde milieu biedt weinig ruimte voor bacteriën en andere organismen die als besmettingen het produkt bedreigen, om zich ten koste van dat produkt te ontwikkelen. De zuivelindustrie gebruikt hiervoor robuuste mengsels van melkzuurbacteriën die als "startercultures" bekend staan. Veel hiervan zijn bij NIZO food research ont-

wikkeld en worden door CSK in Ede geproduceerd. Daarmee wordt de yoghurt- of kaasbereiding ingezet. De melkeiwitten, met name caseïne, worden door de bacteriën gesplitst in brokstukken, de peptiden en aminozuren, die als voedsel dienen voor de veeleisende melkzuurbacteriën.

Melkzuurbacteriën



De stofwisselingsprodukten die melkzuurbacteriën (of andere micro-organismen, zoals de schimmels die bij de bereiding van bepaalde kaassoorten worden toegepast) tijdens hun groei in de melk maken, geven aan de produkten een bijzondere smaak en kunnen ook de voedingswaarde verhogen. Sommige melkzuurbacteriën kunnen bijvoorbeeld bepaalde vitamines produceren, en andere weer bioactieve peptiden die het immuunsysteem kunnen stimuleren. Verschillende melkzuurbacteriën maken bacteriocines, kleine eiwitten zoals nisine, die als conserveermiddel in voeding gebruikt kunnen worden.

- Ons onderzoek richt zich nu op de volgende vragen:
- Wat zijn de biosyntheseroutes voor de diverse producten?
 - Hoe zijn deze routes gekoppeld en gereguleerd in het totale cellulaire metabolisme?
 - Waarom maakt de éne *Lactococcus*-stam bepaalde smaakstoffen wel, terwijl een zeer verwante stam dat niet doet?
 - Hoe kunnen we deze processen sturen in de richting van gewenste producten?
 - Kunnen we stammen snel screenen en selecteren op basis van gewenste eigenschappen?

Deze vragen kunnen hopelijk beantwoord worden als we inzicht hebben in totale genomen, in alle genen en hun gecodeerde eiwitten.

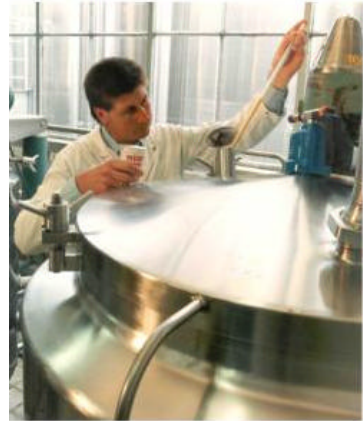
Genoomonderzoek aan melkzuurbacteriën, bacilli en clostridia

De laatste jaren is er een snelle toename geweest van genomische informatie van voedselmicro-organismen, met name gram-positieve bacteriën (Tabel 6). Onlangs zijn de volledige genomesequenties van *Bacillus subtilis*, *Lactococcus lactis* stam IL1403 en *Clostridium acetobutylicum* beschikbaar gekomen. In de loop van 2001 zullen genomesequenties van diverse andere gram-positieve bacteriën, waaronder *Streptococcus thermophilus* (België) en diverse *Lactobacillus*-soorten (o.a VS, Zwitserland, Nederland) voltooid zijn en deels openbaar gemaakt worden (zie Tabel 6).

Tabel 6. Voorbeelden van genomisequencing-projecten van voedsel-relevante “good

Bacterie	Stam	Research-eenheden	Gereed
<i>Bacillus subtilis</i>	168	EU, Japan	1998
<i>Clostridium acetobutylicum</i>	ATCC824	Genome Therapeutics, USA	1999
<i>Lactococcus lactis</i>	IL1403	INRA, FR	2001
<i>Lactococcus lactis</i>	MG1363	RUG, NL; UC Cork, IRL	2001?
<i>Lactobacillus plantarum</i>	WCFS1	WCFS, NL	2001
<i>Lactobacillus acidophilus</i>	ATCC700396	Env. Biotech. Inst., Ca, USA	2001?
<i>Lactobacillus bulgaricus</i>	?	Genoscope, FR	2001?
<i>Lactobacillus helveticus</i>	?	Univ. Wisconsin, USA	2002?
<i>Lactobacillus rhamnosus</i>	?	NZ Dairy Research, NZ	2001?
<i>Lactobacillus johnsonii</i>	NCC533	Nestlé, CH	2000
<i>Streptococcus thermophilus</i>	LMG 18311	UC Louvain, B	2001?
<i>Bifidobacterium sp.</i>	?	UC Cork, IRL	2001?

De omvang van de internationale activiteiten op dit gebied onderstreept het belang van deze benaderingen en de druk op de Nederlandse concurrentiepositie. Nederland heeft een heel sterke positie opgebouwd met betrekking tot de kennis, verbetering en exploitatie van melkzuurbacteriën, door jarenlang onderzoek bij NIZO food research, de Universiteit Groningen, TNO Voeding, en Wageningen Universiteit en Research Centrum. Onlangs is de volledige DNA-volgorde van de ruim 3,3 miljoen basen in het chromosoom van *Lactobacillus plantarum* WCFS1 in Nederland bepaald. De sequentiebepaling is uitgevoerd door het bedrijfje Greenomics in Wageningen, in opdracht van onze projectgroep onder leiding van Michiel Kleerebezem bij het Wageningen Center for Food Sciences (WCFS). Momenteel zijn we druk bezig met de annotatie (= functietoekenning) van de ruim 3500 voorspelde genen, gebruikmakend van de bio-informatica-tools en -databanken bij het CMBI van deze universiteit.



Functionele genomanalyse van deze melkzuurbacteriën kan nu snel gestalte krijgen, omdat er uitstekende genetische tools beschikbaar zijn van deze organismen en uitgebreide kennis over genregulatie, metabolische routes, stress-respons, proteolytische systemen, enz.

Plasmiden

Een bijzonderheid van melkzuurbacteriën is dat sommige genen die coderen voor unieke eigenschappen niet alleen op het chromosoom liggen, maar ook op plasmiden, kleinere stukken circulair DNA. Plasmiden kunnen dikwijls overgedragen worden op andere stammen. Kennis over deze plasmiden is nog erg fragmentarisch. Vooral in industriële *Lactococcus lactis*-stammen vormen plasmiden een normale component van het genoom. Industriële stammen

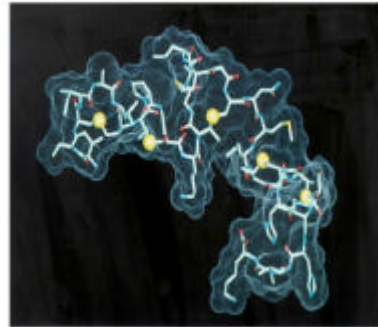


ontlenen vele van hun essentiële en unieke karakteristieken aan diverse stam-specifieke plasmiden. *L. lactis*-plasmiden variëren in grootte van 2 tot >130 kb. De meeste *L. lactis*-stammen hebben 4-7 verschillende plasmiden, maar sommige stammen hebben >10 verschillende plasmiden. Enkele belangrijke plasmide-gecodeerde eigenschappen van *L. lactis* zijn:

- lactoseomzetting en zuurvorming;
- citraatomzetting;
- bacteriocinevorming en -resistentie;
- eiwitafbraak;
- bacteriofaagresistentie;
- exopolysaccharidevorming.

Genoomonderzoek aan melkzuurbacteriën zal zich dus moeten richten op zowel chromosomaal als plasmide-DNA.

Nisine



Vergelijkend genoomonderzoek

Vergelijkend genoomonderzoek van melkzuurbacteriën kan voor de voedingsmiddelenindustrie van groot belang zijn. In principe kan de DNA-volgorde van een bacterieel genoom nu binnen enkele weken volledig worden bepaald, zodat ieder nieuw project met een industrieel micro-organisme zou kunnen beginnen met de genoomsequentie als uitgangspunt. Maar de verscheidenheid aan stammen en soorten melkzuurbacteriën is zo groot dat er bijzondere behoefte bestaat aan de voorspelling van eigenschappen gebaseerd op vergelijkende studies van niet alleen genomen en plasmiden, maar in het bijzonder ook de eiwitten en metabolieten.

Er is nog weinig vergelijkend genoomonderzoek van gram-positieve bacteriën uitgevoerd in de wereld, zodat hierin een kans ligt voor Nederland. Aangezien CMBI, RUG, WCFS en NIZO food research door allianties toegang hebben tot lopende genoomsequencing-resultaten van enkele lactokokken, lactobacillen en streptokokken, kan hiermee een voor-sprong verkregen worden op onderzoeksgroepen in andere landen.

GENOMEN MELKEN

Resultaten van bacterieel genoomonderzoek

Wat hebben we tot nog toe geleerd van hele bacteriële genomen? In het algemeen hebben we meer inzicht, maar vooral overzicht gekregen van:

- het volledige potentieel aan genen en eiwitten in een organisme;
- alternatieve stofwisselingsroutes (en enzymen) naast de klassieke routes;
- eiwit-eiwit-interacties in cellen;
- toxiciteit- en virulentie-eiwitten (targets voor drugs) van ziekteverwekkers;
- de diversiteit en evolutie van bacteriën.

Meer in het bijzonder hebben we geleerd dat:

- duplicaties van genen toenemen met de genoomgrootte (in *E. coli* en *Bacillus* al ca. 50%);
- functioneel gekoppelde genen bij elkaar blijven (in operons, clusters) tijdens evolutie (25);

- horizontale genoverdracht tussen bacteriën veel vaker voorkomt dan gedacht (bijvoorbeeld drug-resistentie, virulentie).

Ronduit verrassend waren de vondsten:

- variaties op één circulair chromosoom mogelijk (35): meerdere chromosomen; lineaire chromosomen;
- hoge plasticiteit van genomen (36): voortdurende veranderingen door inserties/deleties/duplicaties (door homologe recombinatie, transposons, integrons, IS-elementen, bacteriofagen);
- veel geconserveerde eiwitten/families met nog onbekende functie (ca. 15%);
- veel unieke eiwitten in iedere bacterie (gemiddeld 25%).

Uitdagingen voor de komende jaren

Algemene uitdagingen in genomics en bio-informatica zullen zijn:

- standaardiseren, reorganiseren en stroomlijnen van data-opslag en -gebruik;;
- voorspellen en bepalen van de functies van de resterende hypothetische eiwitten;
- voorspellen van de 3D-structuren van zoveel mogelijk gecodeerde eiwitten;
- voorspellen en bepalen van eiwit-eiwitinteracties;
- koppelen van *high-throughput* experimentele biologie aan *high-speed computational biology*.

Uitdagingen die meer toegespitst zijn op industriële applicaties:

- in kaart brengen van volledige genomen van voedsel-relevante bacteriën;
- genexpressiepatronen bepalen onder industriële condities (*transcriptomics*);
- eiwitprofielen bepalen onder verschillende industriële condities (*proteomics*).

Toepassingsmogelijkheden

Welke toepassingen kunnen we verwachten uit genomen van voedsel-relevante bacteriën?

Wat betreft de ongewenste bacteriën zullen er methoden worden ontwikkeld voor:

- Snelle en betrouwbare identificatie van soorten en stammen, op basis van relevante genen en eiwitten die verantwoordelijk zijn voor:
 - tolerantie tegen stress-factoren, zoals zuur, hitte, uitdrogen, voedsellimitatie, zuurstof en antibiotica;
 - ontwikkeling van immuniteit en resistentie tegen antibiotica;
 - toxische factoren (en hun biosyntheseroutes);
 - cel-celinteracties tussen bacterie en gastheer (hechting en kolonisatie).
- Effectieve *tailor-made* bestrijding van specifieke bacteriën.

Wat betreft de nuttige bacteriën zal dit in het algemeen kunnen leiden tot:

- fingerprinting van soorten en stammen;
- voorspelling van functionele eigenschappen;
- identificatie en toepassing van nieuwe metabolieten (bijvoorbeeld bacteriocines);

- sturing van metabolisme voor verbeterde productie van ingrediënten (zoals smaakcomponenten, aminozuren, vitamines, polysacchariden, bacteriocines, antibiotica of andere farmaceutica).

Voor startercultures zal genoomkennis kunnen worden gebruikt voor:

- gerichte selectie van melkzuurbacteriestammen met gewenste eigenschappen, bijvoorbeeld:
 - specifieke smaakvorming;
 - bacteriofaagresistentie;
 - bacteriocineproductie;
 - eiwitafbreekende activiteit;
 - groei op bijzondere koolstofbronnen;
 - resistentie tegen zware metalen;
- verbetering van de stabiliteit en activiteit van starterbacteriën, door het pre-conditioneren van cultures;
- monitoring van startergedrag onder industriële condities (m.b.v. DNA-arrays, proteomics);
- introductie van plasmide-gecodeerde eigenschappen in industriële stammen.

Kennis van genomen van zogenaamde "gezonde" bacteriën zou kunnen worden gebruikt voor:

- gerichte selectie van stammen met gewenste eigenschappen, zoals:
 - specifieke gezondheidsbevorderende verbindingen,
 - overleving en hechting in de darm,
 - verdringing van schadelijke bacteriën (bijv. *Lactobacillus* kan hechting van *Staphylococcus aureus* tegengaan).
- Onderbouwing van gezondheidsclaims.

Bovenstaande voorbeelden benadrukken dat genoomonderzoek of *genomics* een heel andere benadering beoogt dan *genetische modificatie*. Genomics is het in kaart brengen van alle genen en eiwitten, en het voorspellen van alle erfelijke eigenschappen van bacteriën die van nature zijn vastgelegd in hun DNA. Genetische modificatie gaat een stap verder, en is gericht op het veranderen van die erfelijke eigenschappen m.b.v. recombinant-DNA-technieken.

FINANCIERING

Wie gaat dit genoom- en bio-informatica-onderzoek in de komende jaren betalen? In de VS wordt de sequentiebepaling van microbiële genomen al jaren lang fors gefinancierd, door o.a. de National Institutes of Health (NIH), National Science Foundation (NSF) en de Department of Energy (DOE). Ook in Engeland (o.a. Wellcome Trust) en Japan wordt er flink geïnvesteerd.

In Nederland kwam de hype van genomics en bio-informatica pas op gang nadat het ministerie van EZ op 9 september 1999 een Technologie-radarworkshop organiseerde door deskundigen uit het hele land bijeen te roepen. Toen werd gevraagd om een strategisch plan voor de BV Nederland op te stellen. Dit landelijke Strategisch Actieplan Genomics is uitgewerkt door Agro-Food en Bio-Medische stuurgroepen, en in mei 2000 aangeboden aan verschillende ministers.

Momenteel ligt een aanvraag voor 500 miljoen aanvullende financiering van de overheid ter beoordeling bij de commissie Wijffels; eind februari is een tussentijds advies uitgebracht aan de minister. Zeker is dat Voeding en Gezondheid een kernthema zal zijn. Maar goedkeuring zal mede afhankelijk zijn van een forse bijdrage van het bedrijfsleven. Inmiddels zijn eind 2000 twee kleinere programma's, van 70 miljoen gulden voor ruim vijf jaar samen, van start gegaan: het IOP Genomics-programma van EZ, en het Biomoleculaire Informatica-programma van NWO.

Strategisch actieplan GENOMICS

Innovatie en versterking van de kennisinfrastructuur in Nederland



ONDERWIJS

In Nederland ontstaat een grote behoefte aan mensen die zijn opgeleid of getraind in genomics of bio-informatica. Zulke mensen zijn er nog weinig. Specialistische cursussen worden sinds een paar jaar gegeven, o.a. bij de universiteiten van Nijmegen, Wageningen, Utrecht, Amsterdam, Eindhoven en Delft. Verschillende universiteiten zijn nu ook begonnen om bio-informatica en genomics op te nemen in de standaardopleidingen. De eerste hoogleraren Bio-informatica zijn recent benoemd, Prof. Gert Vriend (directeur CMBI) bij de Katholieke Universiteit Nijmegen, en Prof. Willem Stiekema (PRI/Greenomics) bij Wageningen Universiteit. Er zullen er meer volgen.

De leeropdracht bij mijn aanstelling bij de subfaculteit Scheikunde van de Katholieke Universiteit Nijmegen luidt: *Structuurchemie van Eiwitten*. Maar dat interpreteer ik nogal breed, dat heeft u uit mijn toelichting vast wel begrepen. Bij het CMBI ben ik een onderzoeksgroep gestart die zich richt op *Bacterial Genomics* (7), de genomen van bacteriën, en natuurlijk op de eiwitten die daarin gecodeerd zijn. We gebruiken bio-informaticatools, *in-silico*-analyse, voor de interpretatie van bacteriële genomen. Studenten en promovendi worden opgeleid in het gebruik van deze methoden, en nemen actief deel aan onderzoeksprojecten die relevant zijn voor de industrie. In de toekomst zal ik ook colleges gaan geven over bacteriële genomen.

CONCLUSIES

In vogelvlucht heb ik u laten zien dat genoomonderzoek tot op heden nog maar het topje van de ijsberg van microbiële diversiteit heeft belicht, en tegelijk heeft aangetoond hoe weinig er bekend is van micro-organismen. In de komende jaren zal de DNA-volgorde van honderden microbiële genomen bepaald worden, en de wetenschappelijke gemeenschap informatie verschaffen over genen coderend voor miljoenen eiwitten. Hiervan zal een groot aantal nieuw zijn en van onbekende functie. Deze nieuwe genen en eiwitten bieden opwindende kansen voor onderzoek en exploitatie. Het nut van vergelijkend genoomonderzoek is al duidelijk bewezen voor het begrijpen van biochemische diversiteit, virulentie en pathogenese, en evolutie van soorten. De bruikbaarheid van vergelijkend genoomonderzoek zal alleen maar toenemen als meer genoom-DNA-volgordes beschikbaar komen. Een van de belangrijkste uitdagingen voor de toekomst zal zijn om technieken te ontwerpen om de functie van genen en gecodeerde eiwitten op grote schaal te voorspellen en bepalen, en om informatie te integreren over de wijze waarop genen en eiwitten op cellulair niveau samenwerken om een levend organisme te vormen en in stand te houden. Het is niet ondenkbaar dat we, door onze kennis van microbiële biologie en biodiversiteit uit te breiden, grote stappen vooruit zullen maken, zowel in de bestrijding van voedselbederfbacteriën als in de identificatie van nuttige eigenschappen van bacteriën die toepassing kunnen vinden in de voedingsmiddelenindustrie.

SLOTWOORDEN

Aan het einde gekomen van mijn rede wil ik graag enkele dankwoorden uitspreken

Allereerst wil ik een woord van dank richten tot het bestuur van de Stichting Nijmeegs Universiteitsfonds (SNUF) dat mij de bijzondere leerstoel Structuurchemie van Eiwitten heeft willen toevertrouwen, het Faculteitsbestuur Natuurwetenschappen, Wiskunde en Informatica en het College van Bestuur van de Katholieke Universiteit Nijmegen dat deze benoeming heeft bekrachtigd.

Enkele collegae-hoogleraren wil ik in het bijzonder bedanken, Cees Hilbers voor het initiëren van de benoemingsprocedure, en Peter Bloemers, voorzitter van de subfaciliteit Scheikunde, voor het loodsen door allerlei universitaire ambtelijke molens. Excuses voor mijn ongeduldigheid hierin, ik heb nu geleerd dat het bedrijfsleven en de universiteit verschillende definities hebben voor "snelle procedures".

De Groningse hoogleraren Erni van Bruggen, Jan Drenth, en Wil Konings wil ik danken voor de gedegen start die jullie me gegeven hebben als student en promovendus in de structuurchemie en biochemie van eiwitten.

Ad Juriaanse, directeur van NIZO food research, jij was degene die mij in de gelegenheid stelde en stimuleerde om me kandidaat te stellen voor deze functie als bijzonder hoogleraar in Nijmegen.

Jan Wouters, voormalig research-directeur van NIZO food research en WCFS, en nu hoogleraar Zuivelkunde in Wageningen. Jan, jij was mijn mentor in de afgelopen 14 jaar bij NIZO. Dank voor je advies, ik heb er enorm veel aan gehad.

Willem de Vos, huidig research-directeur van WCFS en hoogleraar Microbiologie in Wageningen. Onze wegen lopen al decennia parallel, en we hebben meer dan een beetje van elkaar

geleerd. Onze nieuwe samenwerking in het WCFS, met bio-informatica-ondersteuning vanuit Nijmegen, belooft veel voor de toekomst.

Collega's van WCFS en NIZO food research, er zijn er teveel om persoonlijk te noemen. Zonder teamwork bereik je niets, ook niet in de wetenschap, en ik ben jullie allen daarvoor zeer dankbaar.

Collega's van CMBI, het Centrum voor Moleculaire en Biomoleculaire Informatica van deze universiteit, het was niet gemakkelijk om iemand in jullie midden te verwelkomen die er niet elke dag is. Al snel ontdekten we dat biochemici en informatici niet dezelfde taal spreken (Windows vs. UNIX!). Maar met enig geduld zijn we al een eind gevorderd. Communicatie zal de sleutel tot succes moeten zijn. Ik verheug me erop om in deze goede sfeer met jullie door te gaan.

Gert Vriend, hoogleraar Bio-informatica en hoofd van dit centrum sinds vorig jaar, je staat voor een zware taak om dit centrum uit te laten groeien tot het beste van Nederland. Ik hoop dat ik een steentje kan bijdragen.

Tenslotte nog enkele woorden aan mijn familie en gezin. Lieve moeder, je bent gelukkig zover hersteld van je zware operatie dat je deze dag kan bijwonen. Ik draag deze oratie dan ook aan jou op. Mijn zoons Mark, Sander en Bart: fantastisch dat jullie er vandaag ook bij zijn ondanks drukke werkzaamheden. Jullie zijn geen wetenschappers geworden, maar jullie hebben wel het kaartlezen van mij geleerd, en kunnen hiermee nu uitstekend je eigen koers in het leven uitzetten. Lieve Joke, zonder jou was ik nergens gekomen. Ik ben tenslotte maar de kaartlezer, en jij hebt altijd het stuur in handen gehad.

Ik heb gezegd.

Thus progress in many fields of inquiry is measured not by mere discoveries (however enumerated) but by the deepening of the questions people ask of nature. It is no scandal that many of the questions now in people's minds are extensions of questions asked by Aristotle and his contemporaries. They have become more interesting questions, and they are more taxing. Quite apart from the clamant demand for more applications of science, there is also not yet an end in sight to the process of inquiry. The problems that remain unsolved are gargantuan. They will occupy our children and their children and on and on for centuries to come, perhaps even for the rest of time.

(Sir John Maddox, ex-editor Nature Magazine, 1998)

REFERENTIES

Algemeen

Trends Guide to Bio-informatics, 1998, Elsevier Science.

Brown T.A., Brown A. (1999) Genomes (Wiley-Liss) ISBN: 0471316180

Maddox, J. (1998) What remains to be discovered: mapping the secrets of the universe, the origins of life and the future of the human race. Simon & Schuster, Inc. ISBN 0-684-86300-6.

Matt Ridley, M. (1999) GENOME: The autobiography of a species in 23 chapters. Fourth Estate Limited, London. ISBN 1-85702-835-X.

The Human Genome. Nature, Feb.15, 2001, special issue.

Specifiek

1. www.anl.gov/OPA/frontiers96/cave.html
2. www.orientteering.org/;
www.xs4all.nl/~klaver80/NOLB/
3. Fleischman RD, Adams MD, White O, etc. (1995) Whole genome random sequencing and assembly of the *Haemophilus influenzae* Rd. Science 269, 496-512.
4. <http://www.tigr.org>
5. <http://www.tigr.org/tdb/>;
<http://wit.integratedgenomics.com/GOLD/>
6. Elizabeth Pennisi (2000) BREAKTHROUGH OF THE YEAR: Genomics comes of age. Science 22 december 2000.
<http://www.sciencemag.org/cgi/content/full/290/5500/2220>
7. <http://www.cmbi.kun.nl>
8. Nucleic Acids Research, Database Issue, vol. 29, January 2001.
9. <http://www.infobiogen.fr/>
10. Bairoch A. and Apweiler R. (2000) The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 2000. Nucleic Acids Res. 28, 45-48
11. Altschul, S.F. et al (1997) Nucleic Acids Research 25, 3389-3402;
<http://www.ncbi.nlm.nih.gov/BLAST>
12. Software tools voor genoom analyse bij TIGR: <http://www.tigr.org/softlab/>

13. Bork, P. (2000) Powers and pitfalls in sequence analysis. Genome Research 10, 398-400.
14. Nelson, K.E. et al (2000) Status of genome projects for non-pathogenic bacteria and archaea. Nature Biotech. 18, 1049-1054.
15. Stuart et al (2000) Comparison of functional annotation schemes for genomes. Funct. Integr. Genomics 1, 56-69.
16. Shilling, C. Letscher, D. and Palsson, B.O. (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. J. Theor. Biol. 203, 285-301.
17. Karp, P.D., Krummenacker, M., Paley, S. and Wagg, J. (1999) Integrated pathway-genome databases and their role in drug discovery. TIBTECH 17, 275-281.
18. <http://ecocyc.PangeaSystems.com/ecocyc/>
19. <http://www.genome.ad.jp/kegg/kegg.html>
20. <http://wit.mcs.anl.gov/WIT2/>
21. Schilling en Palsson, (2000) Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. J.Theor. Biol. 203, 249-283; Schilling et al, 2000
22. Enright A. et al. (1999) Protein interaction maps for complete genomes based on gene fusion events. Nature 402, 86-90
23. J.-C. Rain et al. (2001) The protein-protein interaction map of *Helicobacter pylori*. Nature 409, 211-215.
24. <http://e-cell.org/>; Tomita M. et al. (1999) E-CELL: Software environment for whole cell simulation. Bio-informatics 15, 72-84
25. Dandekar, T., Snel, B., Huynen, M. and Bork, P. (1998) Conservation of gene order: a fingerprint of proteins that physically interact. Trends Biochem Sci. 23, 324-328; Marcotte E.M. et al. (1999) Detecting protein function and protein-protein interactions from genome sequences. Science 285, 751-753; Marcotte E.M. et al. (1999) A combined algorithm for genome-wide prediction of protein function. Nature 402, 83-86. Marcotte, E.M. (2000) Computational genetics: finding protein function by nonhomology methods. Curr. Opinion

- Structural Biol.* 10, 359-365. Overbeek, R., Fonstein, M., D-Souza, M., Pusch, G.D. and Maltsev, N. (1999) The use of gene clusters to infer functional coupling. *Proc. Natl. Acad. Sci. USA* 96, 2896-2901.
26. Snel, B. et al (2000) STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res.* 28, 3442-3444. <http://www.bork.embl-heidelberg.de/STRING>
 27. Huynen, M.A. and Bork, P. (1998) Measuring genome evolution. *Proc. Natl. Acad. Sci. USA* 95, 5849-5856
 28. Snel B. et al. (1999) Genome phylogeny based on gene content. *Nature Genetics* 21, 108-110
 29. www.jgi.doe.gov/tempweb/JGI_microbial/html/index.html
 30. www.science.doe.gov/ober/microbial.html; www.ornl.gov/microbialgenomes
 31. <http://microbialcellproject.org>
 32. Perna, N.T. et al (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409, 529-533.
 33. Bionieuws 2, pag. 4, 3 februari 2001.
 34. Konings, W.N., Kok, J. Kuipers, O.P. and Poolman, B. (2000) Lactic acid bacteria: the bugs of the new millennium. *Curr Opin Microbiol.* 3, 276-82.
 35. Heidelberg, J.F. et al (2000) DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* 406, 477-483.
 36. Brunder, W. and Karch, H. (2000) Genome plasticity in Enterobacteriaceae. *Int. J. Med. Microbiol.* 290, 153-165.